JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 65B, No. 2, April–June 1961

Some Properties of the Empirical Distribution Function of a Random Process

M. M. Siddiqui

(January 24, 1961)

Let $\{X(t)\}$ be a continuous-time, continuous-in-the-mean, real, strictly stationary random process. If x is a given number, denote by $p_T(x)$ the proportion of the time $X(t)$ exceeds x, $0 \leq t \leq T$. The covariance of $p_T(x_1)$ and $p_T(x_2)$ is obtained. An approximate solution for the variance of a sample quantile is also given. The general results are specialized for Gaussian and χ^2 processes.

1. Introduction

The proportion-of-time distribution plays a very important role in the statistical analysis of time series data. The practice is to plot $p(x)$, the proportion of the time the record, $X(t)$, exceeds x, against x [e.g., 1,2].[1] In general, if $X(t)$ is a wide-sense stationary process and a sample over $0 \leqq t \leqq T$ is available, an estimate, $p_T(x)$, of the probability that $X(t) \geqq x$ is obtained from the sample. In this paper we will derive the covariance of $p_T(x_1)$ and $p_T(x_2)$. An approximate solution for the variance of a sample quantile will also be given. The general results will be specialized for Gaussian and χ^2 processes. The Rayleigh process is a special case of the χ^2 process; however, due to its wide applications, the result for this particular case will also be stated.

2. Some Properties of Time Averages

Let $\{X_1(t), X_2(t)\}$, $t \epsilon R^{(1)}$, be a wide-sense stationary, real, continuous-parameter, continuous-in-the-mean, vector process. t will be called time and $R^{(1)}$ will be taken to be a real line.

For $i=1, 2$, and for all t and s, let

$$\left.\begin{aligned} &EX_i(t)=m_i, \\ &EX_i(t)X_i(t+s)-m_i^2=R_i(s), \\ &EX_1(t)X_2(t+s)-m_1m_2=R_{12}(s), \end{aligned}\right\} \tag{2.1}$$

where EX denotes the mathematical expectation of the random variable X. It will be assumed that all quantities in (2.1) are finite.

When statements are made which apply to either of the processes, or when the properties of a single process are considered, the process will be referred to as $\{X(t)\}$, and the subscripts will be omitted, e.g., $EX(t)=m$, $\quad EX(t)X(t+s)-m^2=R(s)$.

Let T be a positive number, N a positive integer, $d=T/N$, and denote $X_i(kd)$ by $X_i(k)$, $R_i(kd)$ by $R_i(k)$, etc., since there will be little danger of confusion. $\{X_1(k), X_2(k)\}$, $k=0, \pm 1, \pm 2, \ldots$, is, then, a wide-sense stationary, real, discrete-parameter, vector process.

We may consider the discrete-time process and extend the results to the continuous-time process by taking the limit as $N \rightarrow \infty$, T fixed; or, we may consider the continuous-time process and obtain results for the discrete-time process by replacing integrals by appropriate summations. In general, we will pursue the latter procedure.

[1] Figures in brackets indicate the literature references at the end of this paper.

Let t_1, t_2 be arbitrary numbers, and T_1, T_2 arbitrary positive numbers. Define

$$\overline{X}_{iT_i}(t_i)=\frac{1}{T_i}\int_{t_i}^{t_i+T_i} X_i(t)\,dt, \qquad i=1,2. \tag{2.2}$$

It will be assumed that the subscripts on X are so chosen that $T_2 \geqq T_1$.

For the discrete-time case, we take k_1, k_2 arbitrary integers; N_1, N_2 arbitrary positive integers, choosing the subscripts on X so that $N_2 \geqq N_1$; and define

$$\overline{X}_{iN_i}(k_i)=\frac{1}{N_i}\sum_{k=k_i+1}^{N_i+k_i} X_i(k), \qquad i=1,2. \tag{2.2'}$$

Now

$$E\overline{X}_{iT_i}(t_i)=m_i, \qquad i=1,2, \tag{2.3}$$

and

$$\operatorname{cov}\{\overline{X}_{1T_1}(t_1), \overline{X}_{2T_2}(t_2)\}=(T_1 T_2)^{-1} E\int_{t_1}^{T_1+t_1}\int_{t_2}^{T_2+t_2}[X_1(t')-m_1]\,[X_2(t)-m_2]\,dt\,dt'$$

$$=(T_1T_2)^{-1}\int_{t_1}^{T_1+t_1}\int_{t_2}^{T_2+t_2}R_{12}(t-t')dt\,dt'. \tag{2.4}$$

The interchange of operations of expectation and integration in (2.3) and (2.4) is justified as $EX_i^2(t)<\infty$, $i=1,2$.

We now make the transformation $u=t-t'$, $v=t'$. The Jacobian of the transformation is unity. The limits of u and v are given by

$$t_2-t_1-T_1 \leqq u \leqq t_2-t_1+T_2,$$

$$\max\,(t_1, t_2-u) \leqq v \leqq \min\,(T_1+t_1, T_2+t_2-u).$$

Hence (2.4) becomes

$$(T_1 T_2)^{-1}\int_{-T_1+s}^{T_2+s}[\min\,(T_1+t_1, T_2+t_2-u)-\max\,(t_1, t_2-u)]\,R_{12}(u)\,du,$$

where $s=t_2-t_1$. Now

$$\min\,(T_1+t_1, T_2+t_2-u)=\begin{cases} T_1+t_1, \text{ when } u \leqq s+T_2-T_1 \\ T_2+t_2-u, \qquad \text{otherwise}; \end{cases}$$

$$\max\,(t_1,t_2-u)=\begin{cases} t_2-u, \text{ when } u\leqq s, \\ t_1, \text{ otherwise.} \end{cases} \tag{2.5}$$

Replacing t_1 by t, we obtain, for arbitrary t and s,

$$\operatorname{cov}\,\{\overline{X}_{1T_1}(t), \overline{X}_{2T_2}(t+s)\}=(T_1T_2)^{-1}\Bigg[\int_{-T_1+s}^{s}(T_1-s+u)R_{12}(u)du+T_1\int_{s}^{T_2-T_1+s}R_{12}(u)du$$

$$+\int_{T_2-T_1+s}^{T_2+s}(T_2+s-u)R_{12}(u)du\Bigg]. \tag{2.6}$$

Remark 1. Setting $X_1(t)=X_2(t)=X(t)$ in (2.6), we obtain

$$\operatorname{cov}\,\{\overline{X}_{T_1}(t), \overline{X}_{T_2}(t+s)\}=(T_1T_2)^{-1}\Bigg[\int_{-T_1+s}^{s}(T_1-s+u)R(u)du$$

$$+T_1\int_{s}^{T_2-T_1+s}R(u)du+\int_{T_2-T_1+s}^{T_2+s}(T_2+s-u)R(u)du\Bigg]. \tag{2.7}$$

Remark 2. If $T_1=T_2=T$,

$$\operatorname{cov}\{\overline{X}_T(t), X_T(t+s)\}=T^{-2}\left[\int_{-T+s}^{s}(T-s+u)R(u)du+\int_{s}^{T+s}(T+s-u)R(u)du\right]. \tag{2.8}$$

Remark 3. Setting $s=0$ in (2.8), we have

$$\operatorname{var}\overline{X}_T(t)=T^{-1}\int_{-T}^{T}\left(1-\frac{|u|}{T}\right)R(u)du. \tag{2.9}$$

From (2.3) to (2.9) we deduce that, for fixed T_1 and T_2, $\{\overline{X}_{1T_1}(t), \overline{X}_{2T_2}(t)\}$ is a wide-sense stationary continuous-time and continuous-in-the-mean process.

For the discrete-time case, we will state only the results corresponding to (2.6) and (2.9). When translating a sum of integrals, such as

$$\int_a^b+\int_b^c$$

into a Riemann sum, we should be careful to note that, whereas b appears twice in the limits of the integrals, the corresponding quantity will appear only once in the sum of two summations. Thus, for arbitrary integers k and s,

$$\operatorname{cov}\{\overline{X}_{1N_1}(k), X_{2N_2}(k+s)\}=(N_1N_2)^{-1}\left[\sum_{u=-N_1+s+1}^{s}(N_1-s+u)R_{12}(u)\right.$$
$$\left.+\sum_{u=s+1}^{N_2-N_1+s}N_1R_{12}(u)+\sum_{u=N_2-N_1+s+1}^{N_2+s-1}(N_2+s-u)R_{12}(u)\right]. \tag{2.6$'$}$$

Similarly,

$$\operatorname{var}\overline{X}_N(k)=\frac{R(0)}{N}+\frac{2}{N}\sum_{u=1}^{N-1}\left(1-\frac{u}{N}\right)R(u). \tag{2.9$'$}$$

The results (2.9) and (2.9′) are well known [3, p. 80].

3. Empirical Distribution Function

It will be assumed further that $\{X(t)\}$ is a strictly stationary process so that the probability distribution of $X(t_1+h), \ldots, X(t_n+h)$ is the same as that of $X(t_1), \ldots, X(t_n)$ for every selection of $n, t_1, \ldots, t_n$, and h.

Define

$$p(t;x)=\begin{cases}1, \text{ if } X(t)\geqq x,\\ 0, \text{ otherwise.}\end{cases} \tag{3.1}$$

Then

$$p_T(x)=\frac{1}{T}\int_0^T p(t;x)dt \tag{3.2}$$

is the proportion of the time x is exceeded by $X(t)$ in the time interval $(0, T)$, i.e., $Tp_T(x)$ is the Lebesgue measure of the set $S_x=\{t:X(t)\geqq x, 0\leqq t\leqq T\}$. We note the following:

(1) $p(t;-\infty)=1$ for all t; hence $p_T(-\infty)=1$;
(2) $p(t;+\infty)=0$ for all t; hence $p_T(+\infty)=0$;
(3) If $x_2>x_1$, then $X(t)\geqq x_2$ implies $X(t)\geqq x_1$.

S_{x_2} is, therefore, a subset of S_{x_1}, and hence

$$p_T(x_2)\leqq p_T(x_1).$$

Properties (1) to (3) imply that $q_T(x)=1-p_T(x)$ is a distribution function. $q_T(x)$ will be called

the *empirical univariate distribution function* of the process $\{X(t)\}$. We will, however, consider $p_T(x)$ rather than $q_T(x)$.

Suppressing x from $p(t; x)$ for the time being, we have

$$\begin{aligned}
&\Pr[p(t_1+h)=1, \ldots, \quad p(t_k+h)=1, \quad p(t_{k+1}+h)=0, \ldots, \quad p(t_n+h)=0] \\
&\quad =\Pr[X(t_1+h)\geqq x, \ldots, \quad X(t_k+h)\geqq x, \quad X(t_{k+1}+h)<x, \ldots, \quad X(t_n+h)<x] \\
&\quad =\Pr[X(t_1)\geqq x, \ldots, \quad X(t_k)\geqq x, \quad X(t_{k+1})<x, \ldots, \quad X(t_n)<x] \\
&\quad =\Pr[p(t_1)=1, \ldots, \quad p(t_k)=1, \quad p(t_{k+1})=0, \ldots, \quad p(t_n)=0],
\end{aligned} \tag{3.3}$$

for every selection of k, n, t_1, ... t_k, . . . , t_n, and h. Hence, for a given x, $\{p(t; x)\}$ is a strictly stationary binomial random process. In fact, it can be shown that, for a given set of values, x_1, . . . , x_m, $\{p(t; x_1), \ldots, p(t; x_m)\}$ is a strictly stationary vector process.

Writing

$$P(x)=\Pr(X(t)\geqq x), \tag{3.4}$$

we have

$$\left.\begin{aligned}
&Ep^r(t; x)=P(x), \quad r=1, 2, \ldots, \\
&\operatorname{var} p(t; x)=P(x)[1-P(x)], \\
&A(s; x)\equiv\operatorname{cov}\{p(t; x), p(t+s; x)\} \\
&\qquad =\Pr\{X(0)\geqq x, X(s)\geqq x\}-P^2(x).
\end{aligned}\right\} \tag{3.5}$$

Let $x_2\geqq x_1$; then

$$\begin{aligned}
A(s; x_1, x_2)&\equiv\operatorname{cov}\{p(t; x_1), p(t+s; x_2)\} \\
&=Ep(t; x_1)p(t+s; x_2)-P(x_1)P(x_2) \\
&=\Pr\{X(t)\geqq x_1, X(t+s)\geqq x_2\}-P(x_1)P(x_2) \\
&=\Pr\{X(0)\geqq x_1, X(s)\geqq x_2\}-P(x_1)P(x_2).
\end{aligned} \tag{3.6}$$

We note the following:

$$Ep_T(x)=P(x)$$

$$A(s; x, x)=A(s; x),$$

$$A(0; x_1, x_2)=P(x_2)[1-P(x_1)].$$

Using (2.6) with $s=0$, $T_1=T_2=T$, we have

$$\left.\begin{aligned}
&C(x_1, x_2)\equiv\operatorname{cov}\{p_T(x_1), p_T(x_2)\}=\frac{1}{T}\int_{-T}^{T}\left(1-\frac{|s|}{T}\right)A(s; x_1, x_2)ds, \\
&C(x, x)=\operatorname{var} p_T(x)=\frac{2}{T}\int_0^T\left(1-\frac{s}{T}\right)A(s; x)ds.
\end{aligned}\right\} \tag{3.7}$$

For the discrete-time case,

$$\left.\begin{aligned}
&p_N(x)=N^{-1}\sum_{k=1}^{N}p(k; x), \\
&\operatorname{cov}\{p_N(x_1), p_N(x_2)\}=N^{-1}P(x_2)[1-P(x_1)]+2N^{-1}\sum_{s=1}^{N-1}(1-s/N)A(s; x_1, x_2), \\
&\operatorname{var} p_N(x)=N^{-1}P(x)[1-P(x)]+2N^{-1}\sum_{s=1}^{N-1}(1-s/N)A(s; x).
\end{aligned}\right\} \tag{3.7$'$}$$

To estimate $A(s; x_1, x_2)$ and $A(s; x)$, define

$$p(t, t+s; x_1, x_2)=\begin{cases} 1 \text{ if } X(t)\geqq x_1,\ X(t+s)\geqq x_2 \\ 0 \text{ otherwise;} \end{cases}$$

and use

$$P_T(s; x_1, x_2)=\frac{1}{T-s}\int_0^{T-s} p(t, t+s; x_1, x_2)dt$$

to estimate $\Pr\{X(t)\geqq x_1,\ X(t+s)\geqq x_2\}$.

4. Variance of a Quantile

Given x, $p_T(x)$ is a random variable. In the preceding section we obtained the variance of $p_T(x)$, and the covariance of $p_T(x_1)$ and $p_T(x_2)$. In this section we will obtain an approximation to the variance of a quantile of the empirical distribution function. We suppose that a number q, $0<q<1$, is specified. If U denotes the number such that

$$p_T(U)=q, \tag{4.1}$$

then U is a random variable. Let ξ be the number such that

$$P(\xi)=q, \tag{4.2}$$

i. e., ξ is the qth quantile of the univariate distribution function of the process $[X(t)]$.

In what follows the primes will denote differentiation with respect to the indicated argument. Furthermore, it will be assumed that $P(\xi)$ and $p_T(\xi)$ have derivatives.

Now

$$P(\xi)=p_T(U)\doteq p_T(\xi)+(U-\xi)p'_T(\xi),$$

where $\doteq$ means "approximately equal to;" hence,

$$p_T(\xi)-P(\xi)\doteq-(U-\xi)p'_T(\xi); \tag{4.3}$$

$$0\doteq-E(U-\xi)[p'_T(\xi)-P'(\xi)]-P'(\xi)E(U-\xi).$$

Now

$$Ep'_T(\xi)=E\lim_{h\to 0}\frac{p_T(\xi+h)-p_T(\xi)}{h}=\lim_{h\to 0}\frac{P(\xi+h)-P(\xi)}{h}=P'(\xi);$$

therefore

$$\text{cov}\ (U,p'_T(\xi)\doteq-P'(\xi)E(U-\xi).$$

As a first approximation we suppose that

$$EU=\xi, \tag{4.4}$$

so that

$$\text{cov}(U,p'_T(\xi))=0. \tag{4.5}$$

As a further approximation suppose that U and $p'_T(\xi)$ are independent. Writing $\delta y=y-Ey$, where y is a random variable, we have

$$\delta p_T(\xi)\doteq-p'_T(\xi)\delta U,$$

$$\text{var}\ p_T(\xi)\doteq\text{var}UEp'^2_T(\xi). \tag{4.6}$$

Now

$$Ep_T'^2(\xi)=E\lim_{h,\,k\to 0}\left(\frac{p_T(\xi+h)-p_T(\xi)}{h}\right)\left(\frac{p_T(\xi+k)-p_T(\xi)}{k}\right)$$

$$=E\lim_{h,\,k\to 0}(hk)^{-1}[\delta p_T(\xi+h)-\delta p_T(\xi)+P(\xi+h)-P(\xi)][\delta p_T(\xi+k)-\delta p_T(\xi)+P(\xi+k)-P(\xi)]$$

$$=P'^2(\xi)+\lim_{h,\,k\to 0}(hk)^{-1}[C(\xi+h,\xi+k)-C(\xi+h,\xi)-C(\xi,\xi+k)+C(\xi,\xi)]$$

$$=P'^2(\xi)+\left.\frac{\partial^2 C(\xi_1,\xi_2)}{\partial\xi_1\partial\xi_2}\right|_{\xi_1=\xi_2=\xi}.$$

$$\operatorname{var} U\doteq\left[P'^2(\xi)+\left.\frac{\partial^2 C(\xi_1,\xi_2)}{\partial\xi_1\partial\xi_2}\right|_{\xi_1=\xi_2=\xi}\right]^{-1}\operatorname{var} p_T(\xi), \tag{4.7}$$

and var $p_T(\xi)$ is given by (3.7). The result (4.7) assumes the existence of various derivatives appearing in the equation. In applying this result it is necessary to look into various approximating assumptions which are used in arriving at (4.7). The approximation (4.5) seems fairly reasonable; but the assumption of independence of U and $p_T(\xi)$ needs careful examination in each individual case.

From (3.7)

$$\frac{\partial^2 C(x_1,x_2)}{\partial x_1\partial x_2}=\frac{1}{T}\frac{\partial^2}{\partial x_1\partial x_2}\int_{-T}^{T}\left(1-\frac{|s|}{T}\right)A(s;x_1,x_2)ds=\frac{1}{T}\int_{-T}^{T}\left(1-\frac{|s|}{T}\right)\frac{\partial^2 A(s;x_1,x_2)}{\partial x_1\partial x_2}\,ds \tag{4.8}$$

in case the interchange of the integration and the partial differentiation is justified. If

$$\int_{-\infty}^{\infty}\left|\frac{\partial^2 A}{\partial x_1\partial x_2}\right|_{x_1=x_2=\xi}ds<\infty,$$

then

$$\left.\frac{\partial^2 C}{\partial x_1\partial x_2}\right|_{x_1=x_2=\xi}=O\left(\frac{1}{T}\right);$$

and therefore

$$\operatorname{var} U\doteq\operatorname{var} p_T(\xi)/P'^2(\xi)[1+O(T^{-1})] \tag{4.9}$$

5. Applications

5.1. Gaussian Process

If $\{X(t)\}$ is Gaussian, write

$$R(0)=\sigma^2\qquad Y(t)=[X(t)-m]/\sigma\qquad \rho(t)=R(t)/\sigma^2.$$

It is known [4, pp. 355–6] that

$$\Pr\{Y(t)\geqq y_1,\ Y(t+s)\geqq y_2\}=(2\pi)^{-1}[1-\rho^2(s)]^{-1/2}\int_{y_1}^{\infty}\int_{y_2}^{\infty}\exp\left\{-\frac{z_1^2-2\rho(s)z_1z_2+z_2^2}{2[1-\rho^2(s)]}\right\}dz_1dz_2$$

$$=P(y_1)P(y_2)+\sum_{j=1}^{\infty}\rho^j(s)\,\frac{H_{j-1}(y_1)H_{j-1}(y_2)e^{-\frac{y_1^2+y_2^2}{2}}}{2\pi j!}$$

$$=P(y_1)P(y_2)+\sum_{j=1}^{\infty}\rho^j(s)\tau_j(y_1)\tau_j(y_2). \tag{5.1}$$

Here

$$P(y)=(2\pi)^{-1/2}\int_y^{\infty}e^{-z^2/2}dz;$$

$H_r(x)$ is the Hermite polynomial of degree r given by

$$H_r(x)=e^{x^2/2}(-d/dx)^r e^{-x^2/2}=x^r-\frac{r(r-1)}{2\,1!}x^{r-2}+\frac{r(r-1)(r-2)(r-3)}{2^2\,2!}x^{r-4}-\dots,$$

and

$$\tau_r(x)=\frac{H_{r-1}(x)\,e^{-x^2/2}}{(2\pi)^{1/2}\,(r!)^{1/2}}$$

is known as the tetrachoric function of order r, and its tables are available [5].

Thus

$$\left.\begin{aligned} A(s;x_1,x_2)&=\sum_{j=1}^{\infty}\rho^j(s)\tau_j\left(\frac{x_1-m}{\sigma}\right)\tau_j\left(\frac{x_2-m}{\sigma}\right),\\ A(s;x)&=\sum_{j=1}^{\infty}\rho^j(s)\tau_j^2\left(\frac{x-m}{\sigma}\right).\end{aligned}\right\}\tag{5.2}$$

Inserting these values in (3.7), we have, since $\rho(-s)=\rho(s)$,

$$\left.\begin{aligned} C(x_1,x_2)&=\frac{2}{T}\int_0^T(1-s/T)\sum_{j=1}^{\infty}\rho^j(s)\tau_j\left(\frac{x_1-m}{\sigma}\right)\tau_j\left(\frac{x_2-m}{\sigma}\right)ds,\\ C(x,x)&=\frac{2}{T}\int_0^T(1-s/T)\sum_{j=1}^{\infty}\rho^j(s)\tau_j^2\left(\frac{x-m}{\sigma}\right)ds.\end{aligned}\right\}\tag{5.3}$$

5.2. Variance of a Quantile

Note that:

$$\frac{d\tau_j(x)}{dx}=(2\pi)^{-1/2}(j!)^{-1/2}d/dx\,[e^{-x^2/2}H_{j-1}(x)]=-(j+1)^{1/2}\tau_{j+1}(x).$$

Differentiating the first equation in (5.3) with respect to x_1 and x_2, we get the formal expression

$$\left.\frac{\partial^2 C}{\partial x_1\,\partial x_2}\right|_{x_1=x_2=\xi}=\frac{2}{\sigma^2 T}\int_0^T(1-s/T)\sum_{j=1}^{\infty}(j+1)\rho^j(s)\,\tau_{j+1}^2\left(\frac{\xi-m}{\sigma}\right)ds.\tag{5.4}$$

If the integral converges, we can evaluate it, and using (4.7) obtain the variance of the sample quantile U.

A special case. If $x=m$,

$$\operatorname{var}p_T(m)=2T^{-1}\int_0^T(1-s/T)\sum_{j=1}^{\infty}\rho^j(s)\,\tau_j^2(0)\,ds.$$

Now

$$\tau_{2j}(0)=0,\qquad j=1,2,\dots,$$

$$\tau_{2j+1}(0)=\frac{(-1)^j(2j)!}{(2\pi)^{1/2}\,2^j(j!)\{(2j+1)!\}^{1/2}},\qquad j=0,1,\dots\;.$$

Hence

$$\operatorname{var}p_T(m)=\frac{1}{\pi T}\sum_{j=0}^{\infty}\frac{(2j)!}{2^{2j}(j!)^2(2j+1)}\int_0^T(1-s/T)\,\rho^{2j+1}(s)\,ds.\tag{5.5}$$

If U is the sample median, then $\xi=m$, and

$$\frac{\partial^2 C}{\partial x_1\,\partial x_2}\Big|_{x_1=x_2=m}=(\pi\sigma^2 T)^{-1}\int_0^T (1-s/T)\sum_{j=1}^{\infty}\frac{(2j)!}{2^{2j}(j!)^2}\rho^{2j}(s)\,ds$$

$$=(\pi\sigma^2 T)^{-1}\int_0^T (1-s/T)\,[\{1-\rho^2(s)\}^{-1/2}-1]\,ds. \tag{5.6}$$

in case the integral converges. We note that, for large value of j

$$\frac{(2j)!}{2^{2j}(j!)^2}\simeq\frac{1}{(\pi j)^{1/2}}.$$

Hence the series (5.5) will always converge. A sufficient condition for the convergence of the integral in (5.6) is that

$$\lim_{s\to 0} s^b\{1-\rho^2(s)\}^{-1/2}$$

exists for some $b, 0<b<1$. If the limit is nonzero for some $b\geq 1$ then the integral diverges.

Example 1. If $\rho(s)=\rho(-s)=e^{-\lambda s}, \lambda>0, s\geqq 0$,

$$\int_0^T (1-s/T)e^{-j\lambda s}ds=\frac{1}{j\lambda}-\frac{1}{j^2\lambda^2 T}+O(T^{-1}e^{-j\lambda T}),\qquad j=1,2,\ldots.$$

Therefore

$$\operatorname{var} p_T(m)=\frac{1}{\pi\lambda T}\sum_{j=0}^{\infty}\frac{(2j)!}{2^{2j}(j!)^2(2j+1)^2}-\frac{1}{\pi\lambda^2 T^2}\sum_{j=0}^{\infty}\frac{(2j)!}{2^{2j}(j!)^2(2j+1)^3}+O(T^{-1}e^{-\lambda T})$$

$$\doteq 0.35/(\lambda T)-0.33/(\lambda T)^2+O(T^{-1}e^{-\lambda T}).$$

Also

$$\frac{\partial^2 C}{\partial x_1\partial x_2}\Big|_{x_1=x_2=m}=\frac{1}{\lambda\pi\sigma^2 T}\sum_{j=1}^{\infty}\frac{(2j)!}{2^{2j}(j!)^2(2j)}-\frac{1}{\pi\lambda^2\sigma^2 T^2}\sum_{j=1}^{\infty}\frac{(2j)!}{2^{2j}(j!)^2(2j)^2}+O(T^{-1}e^{-\lambda T})$$

$$=\frac{ln 2}{\pi\lambda\sigma^2 T}-\frac{0.17}{\pi\lambda^2\sigma^2 T^2}+O(T^{-1}e^{-\lambda T}).$$

Hence, if U is the sample median,

$$\operatorname{var} U\simeq\left[\frac{1}{2\pi\sigma^2}+\frac{ln 2}{\pi\lambda\sigma^2 T}-\frac{0.17}{\pi\lambda^2\sigma^2 T^2}\right]^{-1}\operatorname{var} p_T(m)$$

$$=2\pi\sigma^2\operatorname{var} p_T(m)\left[1+\frac{2\,ln 2}{\lambda T}-\frac{0.34}{\lambda^2 T^2}\right]^{-1}.$$

Example 2. If $\rho(s)=e^{-\frac{s^2}{2a^2}}$, a similar calculation shows that

$$\operatorname{var} p_T(m)=0.47a/T-0.35a^2/T^2+O\left(T^{-1}e^{-\frac{T^2}{2a^2}}\right).$$

The integral (5.6) does not converge and we fail in evaluating the variance of the sample median by this method. In fact, an examination of the assumptions leading to the equation (4.7) shows that, in this case, $Ep_T'^2(\xi)$ does not exist so that $p_T(\xi)$ is not differentiable.

Example 3. Let $\rho(s)=\cos(\pi s/k)$. Instead of using (5.5) to evaluate $\operatorname{var} p_T(m)$, we proceed somewhat differently. Note that [6, p. 290]

$$\Pr\{Y(t)\geqq 0, Y(t+s)\geqq 0\}=\frac{1}{2}-\frac{1}{2\pi}\cos^{-1}\rho(s),\qquad 0\leqq\cos^{-1}\rho(s)\leqq\pi.$$

Hence

$$\operatorname{var} p_T(m)=\frac{1}{2T}\int_0^T\left(1-\frac{s}{T}\right)ds-\frac{1}{\pi T}\int_0^T\left(1-\frac{s}{T}\right)\cos^{-1}\rho(s)ds.$$

Now

$$\cos^{-1}\rho(s)=\begin{cases}\frac{\pi}{k}(s-2jk), & \text{if } 2jk\leqq s<(2j+1)k, \quad j=0,1,\ldots,\\ -\frac{\pi}{k}(s-2jk), & \text{if } (2j-1)k\leqq s<2jk, \quad j=1,2,\ldots.\end{cases}$$

Let $T=lk+T_1$, $l\geqq 0$, $0\leqq T_1<k$. We have

$$\operatorname{var} p_T(m)=\frac{1}{4}+\sum_{j=0}^{l-1}I_j-\frac{1}{kT}\int_{T-T_1}^{T}\left(1-\frac{s}{T}\right)\cos^{-1}\rho(s)ds,$$

where

$$I_{2j}=-\frac{1}{kT}\int_{2jk}^{(2j+1)k}\left(1-\frac{s}{T}\right)(s-2jk)ds=-\frac{k}{2T}+\frac{k^2}{3T^2}+\frac{jk^2}{T^2},$$

$$I_{2j-1}=\frac{1}{kT}\int_{(2j-1)k}^{2jk}\left(1-\frac{s}{T}\right)(s-2jk)ds=-\frac{k}{2T}-\frac{k^2}{3T^2}+\frac{jk^2}{T^2}.$$

Thus, after some simplification, we obtain

$$\operatorname{var} p_T(m)=\begin{cases}\frac{T_1^2}{12T^2}\left(3-\frac{2T_1}{k}\right), & \text{if } l \text{ is even}\\ \frac{1}{12T^2}\left(k^2-3T_1^2+\frac{2T_1^3}{k}\right) & \text{if } l \text{ is odd.}\end{cases}$$

Obviously $\operatorname{var} p_T(m)=0$ whenever $T_1=0$, l even, i.e., whenever T is a multiple of $2k$, the period of the process, and also $\operatorname{var} p_T(m)\to 0$ as $T\to\infty$.

5.3. χ^2 Process

Let $\{X_i(t)\}$, $i=1, 2, \ldots, n$, be independent stationary Gaussian processes with the same variance, σ^2, and the same autocorrelation function

$$\rho(s)=[EX_i(t)X_i(t+s)-m_i^2]/\sigma^2.$$

Define

$$Y(t)=\sum_{i=1}^{n}\sigma^{-2}\{X_i(t)-m_i\}^2.$$

$\{Y(t)\}$ will be called a χ^2 process as the univariate distribution of $Y(t)$ is a χ^2 distribution with n degrees of freedom which has the density function

$$g(z;n/2)=[2^{n/2}\Gamma(n/2)]^{-1}e^{-z/2}z^{n/2-1}, \text{ for } z>0,$$

$$=0 \text{ otherwise.}$$

It is easy to verify that $\{Y(t)\}$ is strictly stationary.

The characteristic function, $\phi(u,v)$, of $[Y(t), Y(t+s)]$ is easily evaluated to be

$$\phi(u, v)=Ee^{iuY(t)+ivY(t+s)}$$

$$=[1-2iu-2iv-4uv(1-\rho^2(s))]^{-n/2}$$

$$=(1-2iu)^{-n/2}(1-2iv)^{-n/2}\left[1+\frac{4\rho^2(s)uv}{(1-2iu)(1-2iv)}\right]^{-n/2}$$

$$=\sum_{j=0}^{\infty}\frac{(-1)^j\Gamma(n/2+j)}{\Gamma(n/2)\Gamma(j+1)}\,\frac{\{2\rho(s)\}^{2j}u^jv^j}{(1-2iu)^{n/2+j}(1-2iv)^{n/2+j}}; \tag{5.7}$$

as

$$\left|\frac{4uv}{(1-2iu)(1-2iv)}\right|=\frac{4|uv|}{[(1+4u^2)(1+4v^2)]^{1/2}}<1,$$

the binomial expansion is uniformly convergent.

For a full discussion of the distribution of $[Y(t), Y(t+s)]$ the reader is referred to [7]. A technique of obtaining the probability density function of $[Y(t), Y(t+s)]$ will be presented here, which is also applicable to other distributions which can be expanded in a series of orthogonal polynomials.

Note the following:

(1)
$$g(z; n/2)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-iuz}}{(1-2iu)^{n/2}}\,du,$$

$$\frac{d^j}{dz^j}\,g(z; n/2)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{(-iu)^je^{-iuz}}{(1-2iu)^{n/2}}\,du,$$

(2) It is easy to verify from the definition of Laguerre polynomials [8, Ch. 5] that, for $\alpha>0$, $j\leqq\alpha-1$,

$$\frac{d^j}{dz^j}\,g(z; a)=g(z; a-j)\,\frac{\Gamma(j+1)\Gamma(a-j)}{2^j\Gamma(a)}\,L_j^{(a-j-1)}(z/2),$$

where

$$L_j^{(a)}(z)=\sum_{k=0}^{j}(-1)^k\binom{a+j}{j-k}\frac{z^k}{k!},$$

is the generalized Laguerre polynomial of degree j.

Hence, if $n\geqq 2-n=1$ may be reduced to the Gaussian case discussed in 5.1—, by the inversion of (5.7). the probability density function, $g(z_1, z_2; n/2)$, of $[Y(t), Y(t+s)]$ is given by

$$g(z_1, z_2; n/2)=\sum_{k=0}^{\infty}\frac{[2\rho(s)]^{2j}\Gamma(n/2+j)}{\Gamma(n/2)\Gamma(j+1)}\,\frac{d^j}{dz_1^j}\,g(z_1; n/2+j)\,\frac{d^j}{dz_2^j}\,g(z_2; n/2+j)$$

$$=g(z_1; n/2)g(z_2; n/2)\sum_{j=0}^{\infty}\frac{\Gamma(j+1)\Gamma(n/2)}{\Gamma(n/2+j)}\,\rho^{2j}(s)L_j^{(n/2-1)}(z_1/2)L_j^{(n/2-1)}(z_2/2). \tag{5.8}$$

Now, using the first equation in (5.8), we obtain

$$\Pr\{Y(t)\geqq y_1, Y(t+s)\geqq y_2\}=\int_{y_1}^{\infty}\int_{y_2}^{\infty}g(z_1; z_2; n/2)\,dz_2dz_1$$

$$=P(y_1)P(y_2)+\sum_{j=1}^{\infty}\frac{[2\rho(s)]^{2j}\Gamma(n/2+j)}{\Gamma(n/2)\,\Gamma(j+1)}\,\frac{d^{j-1}}{dy_1^{j-1}}g(y_1; n/2+j)\,\frac{d^{j-1}}{dy_2^{j-1}}g(y_2; n/2+j)$$

$$=P(y_1)P(y_2)+g(y_1; n/2+1)g(y_2; n/2+1)\sum_{i=1}^{\infty}\frac{2n\Gamma(j)\Gamma(n/2+1)\rho^{2j}(s)}{j\Gamma(n/2+j)}$$
$$L_{j-1}^{(n/2)}(y_1/2)L_{j-1}^{(n/2)}(y_2/2), \tag{5.9}$$

where

$$P(y)=\int_y^\infty g(z;\, n/2)dz. \tag{5.10}$$

From (3.6) and (5.9), the covariance of $p(t;y_1)$ and $p(t+s;y_2)$,

$$A(s;y_1,y_2)=g(y_1;n/2+1)g(y_2;n/2+1)\sum_{j=1}^{\infty}\frac{2n\Gamma(j)\,\Gamma(n/2+1)\rho^{2j}(s)}{j\Gamma(n/2+j)}L_{j-1}^{(n/2)}(y_1/2)\,L_{j-1}^{(n/2)}(y_2/2).$$

Inserting this value in (3.7), we obtain an expression for the covariance of $p_T(y_1)$ and $p_T(y_2)$. The variance is given by setting $y_1=y_2=y$.

A special case. $n=2$ gives a χ^2 process with two degrees of freedom, i.e., the Rayleigh process. We find

$$\text{cov}\,\{p_T(y_1),p_T(y_2)\}=\frac{8}{T}g(y_1;2)g(y_2;2)\sum_{j=1}^{\infty}\frac{1}{j^2}L_{j-1}^{(1)}(y_1/2)L_{j-1}^{(1)}(y_2/2)\int_0^T\left(1-\frac{s}{T}\right)\rho^{2j}(s)ds. \tag{5.11}$$

The author thanks Dr. E. L. Crow for his helpful discussions during the preparation of this paper.

6. References

[1] A. D. Watt and R. W. Plush, Measured distributions of the instantaneous envelope amplitude and instantaneous frequency of carriers plus thermal and atmospheric noise, Statistical Methods in Radio Wave Propagation, pp. 233–247 (Pergamon Press, London, England, 1960).

[2] Raymond E. McGavin and Leo J. Maloney, Study at 1046 megacycles per second of the reflection coefficient of irregular terrain at grazing angles, J. Research NBS **63D,** 235–248 (1959).

[3] W. B. Davenport and W. L. Root, An Introduction to the Theory of Random Signals and Noises (McGraw-Hill Book Co., Inc., New York, N.Y., 1958).

[4] M. G. Kendall, The Advanced Theory of Statistics, I (Charles Griffin and Co., London, England, 1948).

[5] K. Pearson, Tables for Statisticians and Biometricians, Parts I and II (Cambridge Univ. Press, Cambridge, England, 1930 and 1931).

[6] Harald Cramér, Mathematical Methods of Statistics (Princeton Univ. Press, Princeton, N.J., 1948).

[7] M. Nakagami, The *m*-distribution—A general formula of intensity distribution of radio fading, Statistical Methods in Radio Wave Propagation, pp. 3–36 (Pergamon Press, London, England, 1960).

[8] Gabor Szegö, Orthogonal Polynomials (Am. Math. Soc. New York, N.Y., 1939).

(Paper 65B2–50)